NUCLEIC ACID DETECTION ASSAY CONTROL GENES

INVENTOR: Uwe SCHERF

FIELD OF THE INVENTION

[0001] The invention relates generally to control genes that maybe utilized for normalizing hybridization and/or amplification reactions, as well as methods of identifying these genes that may be used in toxicology studies and in analyzing gene expression data sets for quality and compatibility with other data sets.

RELATED APPLICATIONS

[0002] This application claims priority under 35 U.S.C. §119(e) to U.S. Provisional Application 60/399,158, filed July 30, 2002, which is herein incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

[0003] Nucleic acid hybridization and other quantitative nucleic acid detection assays are routinely used in medical and biotechnological research and development, diagnostic testing, drug development and forensics. Such technologies have been used to identify genes which are up- or down-regulated in various disease or physiological states, to analyze the roles of the members of cellular signaling cascades and to identify drugable targets for various disease and pathology states.

[0004] Examples of technologies commonly used for the detection and/or quantification of nucleic acids include Northern blotting (Krumlauf (1994), *Mol Biotechnol* 2: 227-242), *in situ* hybridization (Parker & Barnes (1999), *Methods Mol Biol* 106: 247-283), RNAse protection assays (Hod (1992), *Biotechniques* 13: 852-854; Saccomanno *et al.* (1992), *Biotechniques* 13: 846-850), microarrays, and reverse transcription polymerase chain reaction (RT-PCR) (see Bustin (2000), *J Mol Endocrin* 25: 169-193).

[0005] The reliability of these nucleic acid detection methods depend on the availability of accurate means for accounting for variations between analyses. For example, variations in hybridization conditions, label intensity, reading and detector efficiency, sample concentration and quality, background effects, and image processing effects each contribute

to signal heterogeneity (Hegde *et al.* (2000), *Biotechniques* 29: 548-562; Berger *et al.* (2000), WO 00/04188). Normalization procedures used to overcome these variations often rely on control hybridizations to housekeeping genes such as P-actin, glyceraldehyde-3 -phosphate dehydrogenase (GADPH), and the transferrin receptor gene (Eickhoff *et al.* (1999), *Nucl Acids Res* 27:e33; Spiess *et al.* (1999), *Biotechniques* 26: 46-50. These methods, however, generally do not provide the signal linearity sufficient to detect small but significant changes in transcription or gene expression (Spiess *et al.*(1999), *Biotechniques* 26:46-50). In addition, the steady state levels of many housekeeping genes are susceptible to alterations in expression levels that are dependent on cell differentiation, nutritional state, specific experimental and stimulation protocols (Eickhoff *et al.* (1999), *Nucl Acids Res* 27:e33; Spiess *et al.* (1999), *Biotechniques* 26:46-50; Hegde *et al.* (2000), *Biotechniques* 29:548-562; and Berger *et al.* (2000), WO 00/04188). Consequently, there exists a need for the identification and use of additional genes that may serve as effective controls in nucleic acid detection assays.

SUMMARY OF THE INVENTION

[0006] The present invention includes methods of identifying at least one gene that is consistently expressed across different cell or tissue types in an organism, comprising: preparing gene expression profiles for different cell or tissue types from the organism; calculating a coefficient of variation for at least one gene in each of the profiles across the different cell or tissue types; and selecting any gene whose coefficient of variation indicates that the gene is consistently expressed across the different cell or tissue types. The coefficient of variation may be less than about 40% and the methods may comprise creating gene expression profiles for about 10, 25, 50, 100 or more different cell or tissue types. The gene expression profiles may be prepared be querying a gene expression database.

[0007] The invention also includes a set of probes comprising at least two probes that specifically hybridize to a control gene identified by the methods of the invention. Such sets of probes may comprise probes that specifically hybridize to at least about 10, 25, 50 or 100 control genes. In some formats, the sets of probes are attached to a solid substrate such as a microarray or chip.

[0008] The invention also includes methods of normalizing the data from a nucleic acid detection assay comprising: detecting the expression level for at least one gene in a nucleic

acid sample; and normalizing the expression of said at least one gene with the detected expression of at least one control gene identified by the method of the invention. The number of control genes used to normalize gene expression data may comprise about 10, 25, 50, 100 or more of the control genes herein identified.

[0009] In another embodiment, the invention includes a set of probes comprising at least two probes that specifically hybridize to a gene of Table 1. The set may comprise at least about 10, 25, 50, 100 or more the control genes of Table 1. The sets of probes may or may not be attached to a solid substrate such as a chip.

[0010] The invention, in another embodiment, includes methods of normalizing the data from a nucleic acid detection assay comprising: detecting the expression level for at least one gene in a nucleic acid sample; and normalizing the expression of said at least one gene with the detected expression of at least one control gene of Table 1. The number of control genes used to normalize gene expression data may comprise about 10, 25, 50, 100, 500 or more of the control genes herein identified.

DETAILED DESCRIPTION

[0011] The present Inventors have identified rat control genes that may be monitored in nucleic acid detection assays and whose expression levels may be used to normalize gene expression data or evaluate the suitability of test data to compare to or to include in a database of like data. Normalization of gene expression data from a cell or tissue sample with the expression level(s) of the identified control genes allows the accurate assessment of the expression level(s) for genes that are differentially regulated between samples, tissues, treatment conditions, etc. These control genes may be used across a broad spectrum of assay formats, but are particularly useful in microarray or hybridization based assay formats.

A. Nucleic Acid Detection Assay Controls

1. Selection of Control Genes

[0012] As used herein, the genes selected by the disclosed methods as well as the rat genes and nucleic acids of Table 1 are referred to as "invariant" or "control genes." Control genes of the invention may be produced by a method comprising preparing gene expression profiles (a representation of the expression level for at least one gene, preferably 10, 25, 50, 100, 500 or more, or, most preferably, nearly all or all expressed genes in a sample) from at least two

(or a variety) of cell or tissue types, or from a set of samples of at least one cell or tissue type in which the set contains normal samples (from healthy animals), disease state samples, toxin-exposed samples, etc., measuring the level of expression for at least one gene in each of the gene expression profiles to produce gene expression data, calculating a coefficient of variation in the expression level from the gene expression data for each gene (%CV) and selecting genes whose coefficient of variation indicates that the gene is consistently expressed at about the same level in the different cell or tissue types. In one embodiment, such genes that are expressed at about the same level, or are invariantly expressed, are those genes that have a coefficient of variation (expressed as a percentage) of less than or equal to about 40%.

[0013] In the methods of the invention, gene expression profiles maybe produced by any means of quantifying gene expression for at least one gene in the tissue or cell sample. In preferred methods, gene expression is quantified by a method selected from the group consisting of a hybridization assay or an amplification assay. Hybridization assays may be based on any assay format that relies on the hybridization of a probe or primer to a nucleic acid molecule in the sample. Such formats include, but are not limited to, differential display formats and microarray hybridization, including microarrays produced in chip format. Amplification assays include, but are not limited to, quantitative PCR, semiquantitative PCR and assays that rely on amplification of nucleic acids subsequent to the hybridization of the nucleic acid to a probe or primer. Such assays include the amplification of nucleic acid molecules from a sample that are bound to a microarray or chip.

[0014] In other circumstances, gene expression profiles may be produced by querying a gene expression database comprising expression results for genes from various cell or tissue samples. The gene expression results in the database may be produced by any available method, such as differential display methods and micro array-based hybridization methods. The gene expression profile is typically produced by the step of querying the database with the identity of a specific cell or tissue type for the genes that are expressed in the cell or tissue type and/or the genes that are differentially regulated compared to a control cell or tissue sample. Available databases include, but are not limited to, the Gene Logic Gene Express™ database, the Gene Expression Omnibus gene expression and hybridization array repository available through NCBI (www.ncbi.nlm.nih.gov/entrez) and the SAGE™ gene expression database.

[0015] In preferred embodiments, the statistical measure referred to herein as the coefficient of variation (%CV) is calculated on a gene by gene basis across a number of samples or across a reference database to find the least variant genes with respect to a number of cell or tissue types or sample treatments.

10016] Further, the statistical methods of the invention are particularly useful for determining the compatibility of a test sample to an entire set of samples, or an existing database derived from those samples. For instance, a %CV value for genes that have been shown to be the most resistant to variability is calculated for all samples within a test group or test database. These %CV values are then compared to those from a standard reference database. Accordingly, a closeness distribution of all individual samples in the test database to the reference database as a whole can be generated to evaluate the compatibility of new samples. The genes identified in Table I show invariant patterns of expression and can be used to assess compatibility and reliability of gene expression experiments and predictive modeling experiments. These genes show low variability both in control groups from many different experiments and in studies of disruptions of gene expression, such as those occurring in disease states. As a result, these genes can be used as an internal standard for comparing gene expression data. Measurements of expression levels of these genes are used to determine the extent of compatibility of data from different sources and the need, or lack thereof, for normalization or further quality control and adjustments. These measurements also provide an internal standard that supplies a reference point for highly disrupted patterns of gene expression. These genes are also of critical importance for determining relative expression if small numbers of markers are used in custom microarrays.

[0017] The cell or tissue samples that are used to prepare gene expression profiles may include any cell or tissue sample available. Such samples include, but are not limited to, tissues removed as surgical samples, diseased or normal tissues, *in vitro* or *in vivo* grown cells, and cell cultures and cells or tissues from animals exposed to an agent such as a toxin. The number of samples that may be used to calculate a coefficient of variation is variable, but may include about 3, 10, 25, 50, 100, 200, 500 or more cell or tissue samples. The cell or tissue samples may be derived from an animal or plant, preferably a mammal, most preferably a rat. In some instances, the cell or tissue samples may be human, canine (dog), mouse or rat in origin.

[0018] In some embodiments of the invention, the coefficient of variation maybe calculated from raw expression data or from data that has been normalized to control for the mechanics of hybridization, such as data normalized or controlled for background noise due to non-specific hybridization. Such data typically includes, but is not limited to, fluorescence readings from microarray based hybridizations, densitometry readings produced from assays that rely on radiological labels to detect and quantify gene expression and data produced from quantitative or semi-quantitative amplification assays.

[0019] The coefficient of variation (CV) is typically calculated by calculating a mean value for the expression level of a given gene across a number of samples and calculating the standard deviation (SD) from that mean. The CV may be calculated by the following equation: CV=SD/Mean and may or may not be presented as a percentile value. Genes with a CV of less than about 40% may be selected as control genes or are considered as genes that are consistently expressed across the different cell or tissue types tested.

[0020] As used herein, "background" refers to signals associated with non-specific binding (cross-hybridization). In addition to cross-hybridization, background may also be produced by intrinsic fluorescence of the hybridization format components themselves.

[0021] "Bind(s) substantially" refers to complementary hybridization between an oligonucleotide probe and a nucleic acid sample and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the nucleic acid sample.

[0022] The phrase "hybridizing specifically to" refers to the binding, duplexing or hybridizing of a molecule substantially to or only to a particular nucleotide sequence or sequences under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA.

2. Preparation of Controls Genes, Probes and Primers

[0023] The control genes listed in Table 1 may be obtained from a variety of natural sources such as organisms, organs, tissues and cells. The sequences of known genes are in the public databases. The GenBank Accession Number corresponding to the Normalization Control Genes can be found in Table 1. The sequences of the genes in GenBank (http://www.ncbi.nlm.nih.gov/) are herein incorporated by reference in their entirety as of the priority date of this application.

[0024] Probes or primers for the nucleic acid detection assays described herein that specifically hybridize to a control gene may be produced by any available means. For instance, probe sequences may be prepared by cleaving DNA molecules produced by standard procedures with commercially available restriction endonucleases or other cleaving agents. Following isolation and purification, these resultant normalization control gene fragments can be used directly, amplified by PCR methods or amplified by replication or expression from a vector.

[0025] Control genes and control gene probes or primers *(i.e.,* synthetic oligonucleotides and polynucleotides) are most easily synthesized by chemical techniques, for example, the phosphoramidite method of Matteucci *et al.* ((1981) *J Am Chem Soc* 103:3185-3191) or using automated synthesis methods using the GenBank sequences disclosed in Table 1. Probes for attachment to microarrays or for use as primers in amplification assays may be produced from the sequences of the genes identified herein using any available software, including, for instance, software available from Molecular Biology Insights, Olympus Optical Co. and Premier Biosoft International.

[0026] In addition, larger nucleic acids can readily be prepared by well known methods, such as synthesis of a group of oligonucleotides that define various modular segments of the normalization control genes and normalization control gene segments, followed by ligation of oligonucleotides to build the complete nucleic acid molecule.

B. Normalization Methods

[0027] Gene expression data produced from the control genes in a given sample or samples may be used to normalize the gene expression data from other genes using any available arithmatic or calculative means. In particular, gene expression data from the control genes in Table 1 are useful to normalize gene expression data for toxicology testing or modeling in an animal model, preferably in a rat. Such methods include, but are not limited, methods of data analysis described by Hegde *et al.* (2000), *Biotechniques* 29:548-562; Winzeller *et al.* (1999), *Meth Enzymol* 306:3-18; Tkatchenko *et al.* (2000), *Biochimica et Biophysica Acta* 1500:17-30; Berger *et al.* (2000), WO 00/04188; Schuchhardt *et al.* (2000), *Nucl Acids Res* 28:e47; Eickhoff *et al.* (1999), *Nucl Acids Res* 27:e33. Micro-array data analysis and image processing software packages and protocols, including normalization methods, are also available from BioDiscovery (http://www.biodiscovery.com), Silicon Graphics

(http://www.sigenetics.com), Spotfire (http://www.spotfire.com), Stanford University (http://rana.Stanford.EDU/software), National Human Genome Research Institute (http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html), TIGR (http://www.tigr.org/softlab), and Affymetrix (affy and maffy packages), among others.

C. Assay or Hybridization Formats

[0028] The control genes of the present invention may be used in any nucleic acid detection assay format, including solution-based and solid support-based assay formats. As used herein, "hybridization assay format(s)" refer to the organization of the oligonucleotide probes relative to the nucleic acid sample. The hybridization assay formats that may be used with the control genes and methods of the present invention include assays where the nucleic acid sample is labeled with one or more detectable labels, assays where the probes are labeled with one or more detectable labels, and assays where the sample or the probes are immobilized. Hybridization assay formats include but are not limited to: Northern blots, Southern blots, dot blots, solution-based assays, branched DNA assays, PCR, RT-PCR, quantitative or semi-quantitative RT-PCR, microarrays and biochips.

[0029] As used herein, "nucleic acid hybridization" simply involves contacting a probe and nucleic acid sample under conditions where the probe and its complementary target can form stable hybrid duplexes through complementary base pairing (see Lockhart *et al.,* (1999) WO 99/32660). The nucleic acids that do not form hybrid duplexes are then washed away leaving the hybridized nucleic acids to be detected, typically through detection of an attached detectable label.

[0030] It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions *(e.g.,* low temperature and/or high salt) hybrid duplexes *(e.g.,* DNA-DNA, RNA-RNA or RNA-DNA) will form even where the annealed sequences are not perfectly complementary. Thus, specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency *(e.g.,* higher temperature or lower salt) successful hybridization requires fewer mismatches. One of skill in the art will appreciate that hybridization conditions may be selected to provide any degree of stringency. In a preferred embodiment, hybridization is performed at low stringency, in this case in 6X SSPE-T at 37°C (0.005% Triton X-100) to ensure hybridization, and then subsequent washes

are performed at higher stringency *(e.g.,* 1X SSPE-T at 37°C) to eliminate mismatched hybrid duplexes. Successive washes may be performed at increasingly higher stringency *(e.g.,* down to as low as 0.25X SSPE-T at 37°C to 50°C until a desired level of hybridization specificity is obtained. Stringency can also be increased by addition of agents such as formamide. Hybridization specificity may be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present *(e.g.,* expression level control, normalization control, mismatch controls, *etc.).*

[0031] As used herein, the term "stringent conditions" refers to conditions under which a probe will hybridize to a complementary control nucleic acid, but with only insubstantial hybridization to other sequences. Stringent conditions are sequence-dependent and will be different under different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

[0032] Typically, stringent conditions will be those in which the salt concentration is at least about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30'C for short probes (e.g., 10 to 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

[0033] In general, there is a tradeoff between hybridization specificity (stringency) and signal intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array may be washed at successively higher stringency solutions and read between each wash. Analysis of the data sets thus produced will reveal a wash stringency above that the hybridization pattern is not appreciably altered and which provides adequate signal for the particular oligonucleotide probes of interest.

[0034] The "percentage of sequence identity" or "sequence identity" is determined by comparing two optimally aligned sequences or subsequences over a comparison window or span, wherein the portion of the polynucleotide sequence in the comparison window may optionally comprise additions or deletions *(i.e.,* gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical residue *(e.g.,* nucleic acid base or amino acid residue) occurs in both sequences to yield the

number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Percentage sequence identity when calculated using the programs GAP or BESTFIT (see below) is calculated using default gap weights. Sequences corresponding to the control genes of the invention may comprise at least about 70% sequence identity to those sequences identified by GenBank Accession Nos. in Table 1, preferably about 75%, 80% or 85% sequence identity, or more preferably, about 90%, 95% or more sequence identity.

[0035] Homology or identity is determined by **BLAST** (Basic Local Alignment Search Tool) analysis using the algorithm employed by the programs blastp, blastn, blastx, tblastn and tblastx (Karlin *et al.* (1990), *Proc Natl Acad Sci* USA 87:2264-2268 and Altschul (1993), *J Mol Evol* 36:290-300, fully incorporated by reference) which are tailored for sequence similarity searching. The approach used by the **BLAST** program is first to consider similar segments between a query sequence and a database sequence, then to evaluate the statistical significance of all matches that are identified and finally to summarize only those matches which satisfy a preselected threshold of significance. For a discussion of basic issues in similarity searching of sequence databases, see Altschul *et al.* (1994), *Nat Genet* 6: 119-129, which is fully incorporated by reference. The search parameters for histogram, descriptions, alignments, expect *(i.e.,* the statistical significance threshold for reporting matches against database sequences), cutoff, matrix and filter are at the default settings. The default scoring matrix used by blastp, blastx, tblastn, and tblastx is the BLOSUM62 matrix (Henikoff *et al.* (1992), *Proc Natl Acad Sci* USA 89:10915-10919, fully incorporated by reference). Four blastn parameters were adjusted as follows: Q=10 (gap creation penalty); R=10 (gap extension penalty); wink=1 (generates word hits at every winkth position along the query); and gapw= 16 (sets the window width within which gapped alignments are generated). The equivalent Blastp parameter settings were Q=9; R=2; wink=l; and gapw=32. A Bestfit comparison between sequences, available in the GCG package version 10.0, uses DNA parameters GAP=50 (gap creation penalty) and LEN=3 (gap extension penalty) and the equivalent settings in protein comparisons are GAP=8 and LEN=2.

[0036] As used herein, a "probe" or "oligonucleotide probe" is defined as a nucleic acid, capable of binding to a nucleic acid sample or complementary control gene nucleic acid through one or more types of chemical bonds, usually through complementary base pairing,

usually through hydrogen bond formation. As used herein, a probe may include natural *(i.e.,* A, G, U, C or T) or modified bases (7-deazaguanosine, inosine, *etc.).* In addition, the bases in probes may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages.

[0037] Probe arrays may contain at least two or more oligonucleotides that are complementary to or hybridize to one or more of the control genes described herein. Such arrays may also contain oligonucleotides that are complementary or hybridize to at least about 2, 3, 5, 7, 10, 50, 100 or more the genes described herein. Any solid surface to which oligonucleotides or nucleic acid sample can be bound, either directly or indirectly, either covalently or non-covalently, can be used. For example, solid supports for various hybridization assay formats can be filters, polyvinyl chloride dishes, silicon or glass based chips, *etc.* Glass-based solid supports, for example, are widely available, as well as associated hybridization protocols (see, *e.g.,* Beattie, WO 95/11755).

[0038] A preferred solid support is a high density array or DNA chip. This contains an oligonucleotide probe of a particular nucleotide sequence at a particular location on the array. Each particular location may contain more than one molecule of the probe, but each molecule within the particular location has an identical sequence. Such particular locations are termed features. There may be, for example, 2, 10, 100, 1000, 10,000, 100,000, 400,000, 1,000,000 or more such features on a single solid support. The solid support, or more specifically, the area wherein the probes are attached, may be on the order of a square centimeter.

1. Dot Blots

[0039] The control genes listed in Table I and methods of the present invention may be utilized in numerous hybridization formats such as dot blots, dipstick, branched DNA sandwich and ELISA assays. Dot blot hybridization assays provide a convenient and efficient method of rapidly analyzing nucleic acid samples in a sensitive manner. Dot blots are generally as sensitive as enzyme-linked immunoassays. Dot blot hybridization analyses are well known in the art and detailed methods of conducting and optimizing these assays are detailed in U.S. Patent Nos. 6,130,042 and 6,129,828, and Tkatchenko *et al.* (2000), *Biochimica et Biophysica Acta* 1500:17-30. Specifically, a labeled or unlabeled nucleic acid sample is denatured, bound to a membrane (*i.e.,* nitrocellulose) and then contacted with

unlabeled or labeled oligonucleotide probes. Buffer and temperature conditions can be adjusted to vary the degree of identity between the oligonucleotide probes and nucleic acid sample necessary for hybridization.

[0040] Several modifications of the basic dot blot hybridization format have been devised. For example, reverse dot blot analyses employ the same strategy as the dot blot method, except that the oligonucleotide probes are bound to the membrane and the nucleic acid sample is applied and hybridized to the bound probes. Similarly, the dot blot hybridization format can be modified to include formats where either the nucleic acid sample or the oligonucleotide probe is applied to microtiter plates, microbeads or other solid substrates.

2. Membrane-Based Formats

[0041] Although each membrane-based format is essentially a variation of the dot blot hybridization format, several types of these formats are preferred. Specifically, the methods of the present invention may be used in Northern and Southern blot hybridization assays. Although the methods of the present invention are generally used in quantitative nucleic acid hybridization assays, these methods may be used in qualitative or semi-quantitative assays such as Southern blots, in order to facilitate comparison of blots. Southern blot hybridization, for example, involves cleavage of either genomic or cDNA with restriction endonucleases followed by separation of the resultant fragments on a polyacrylamide or agarose gel and transfer of the nucleic acid fragments to a membrane filter. Labeled oligonucleotide probes are then hybridized to the membrane-bound nucleic acid fragments. In addition, intact cDNA molecules may also be used, separated by electrophoresis, transferred to a membrane and analyzed by hybridization to labeled probes. Northern analyses, similarly, are conducted on nucleic acids, either intact or fragmented, that are bound to a membrane. The nucleic acids in Northern analyses, however, are generally RNA.

3. Arrays

[0042] Any microarray platform or technology maybe used to produce gene expression data that may be normalized with the control genes and methods of the invention. Oligonucleotide probe arrays can be made and used according to any techniques known in the art (see for example, Lockhart *et al.,* (1996), *Nat Biotechnol* 14: 1675-1680; McGall *et al.* (1996), *Proc Natl Acad Sci USA* 93:13555-13460). Such probe arrays may contain at least

one or more oligonucleotides that are complementary to or hybridize to one or more of the nucleic acids of the nucleic acid sample and/or the control genes of Table 1. Such arrays may also contain oligonucleotides that are complementary or hybridize to at least about 2, 3, 5, 7, 10, 25, 50, 100, 500 or more of the control genes listed in Table 1.

[0043] Control oligonucleotide probes of the invention are preferably of sufficient length to specifically hybridize only to appropriate, complementary genes or transcripts. Typically the oligonucleotide probes will be at least about 10, 12, 14, 16, 18, 20 or 25 nucleotides in length. In some cases longer probes of at least 30, 40, or 50 nucleotides will be desirable. The oligonucleotide probes of high density array chips include oligonucleotides that range from about 5 to about 45, or 5 to about 500 nucleotides, more preferably from about 10 to about 40 nucleotides, and most preferably from about 15 to about 40 nucleotides in length. In other particularly preferred embodiments, the probes are 20 or 25 nucleotides in length. In another preferred embodiment, probes are double- or single-stranded DNA sequences. The oligonucleotide probes are capable of specifically hybridizing to the control gene nucleic acids in a sample.

[0044] One of skill in the art will appreciate that an enormous number of array designs comprising control probes of the invention are suitable for the practice of this invention. The high density array will typically include a number of probes that specifically hybridize to each control gene nucleic acid, *e.g.* mRNA or cRNA (see WO 99/32660 for methods of producing probes for a given gene or genes). Assays and methods comprising control probes of the invention may utilize available formats to simultaneously screen at least about 100, preferably about 1000, more preferably about 10,000 and most preferably about 500,000 or 1,000,000 different nucleic acid hybridizations.

[0045] The methods and control genes of this invention may also be used to normalize gene expression data produced using commercially available oligonucleotide arrays that contain or are modified to contain control gene probes of the invention. A preferred oligonucleotide array may be selected from the Affymetrix, Inc. GeneChip® series of arrays which include the Human Genome Focus Array, Human Genome U133 Set, Human Genome U95 Set, HuGeneFL Array, Human Cancer Array, HuSNP Mapping Array, GenFlex Tag Array, p53 Assay Array, CYP450 Assay Array, Rat Genome U34 Set, Rat Neurobiology U34 Array, Rat Toxicology U34 Array, Murine Genome U74v2, Murine 11K Set, Yeast Genome S98 Array, *E. coli* Antisense Genome Array, *E. coli* Genome Array (Sense), *Arabidopsis* ATH1 Genome

Array, *Arabidopsis* Genome Array, *P. aeruginosa* Genome Array and *B. subtilis* Genome Array. In another embodiment, an oligonucleotide array may be selected from the Motorola Life Sciences and Amersham Pharmaceuticals CodeLink™ Bioarray System microarrays, including the UniSet Human 20K I, Uniset Human I, ADME-Rat, UniSet Rat I and UniSet Mouse I, or from the Motorola Life Sciences eSensor™ series of microarrays.

4. RT-PCR

[0046] The control genes and methods of the invention may be used in any type of polymerase chain reaction. A preferred PCR format is reverse transciptase polymerase chain reaction (RT-PCR), an *in vitro* method for enzymatically amplifying defined sequences of RNA (Rappolee *et al.* (1988), *Science* 241: 708-712) permitting the analysis of different samples from as little as one cell in the same experiment (see "RT-PCR: The Basics," Ambion, www.ambion.com/techlib/basics/rtpcr/index.html; PCR, M.J. McPherson and S.G. Møller, BIOS Scientific Publishers, Oxfordshire, England, 2000; and PCR Primer: A Laboratory Manual, Dieffenbach *et al.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1995, for review). One of ordinary skill in the art may appreciate the enormous number of variations in RT-PCR platforms that are suitable for the practice of the invention, including complex variations aimed at increasing sensitivity such as semi-nested (Wasserman *et al.* (1999), *Mol Diag* 4:21-28), nested (Israeli *et al.* (1994), *Cancer Res* 54:6303-6310; Soeth *et al.* (1996), *Int J Cancer* 69:278-282), and even three-step nested (Funaki *et al.* (1997), *Life Sci* 60:643-652; Funaki *et al.* (1998), *Brit J Cancer* 77:1327-1332).

[0047] In one embodiment of the invention, separate enzymes are used for reverse transcription and PCR amplification. Two commonly used reverse transcriptases, for example, are avian myeloblastosis virus and Moloney murine leukaemia virus. For amplification, a number of thermostable DNA-dependent DNA polymerases are currently available, although they differ in processivity, fidelity, thermal stability and ability to read modified triphosphates such as deoxyuridine and deoxyinosine in the template strand (Adams *et al.* (1994), *Bioorg Med Chem* 2:659-667; Perler *et al.* (1996), *Adv Prot Chem* 48:377-435). The most commonly used enzyme, Taq DNA polymerase, has a 5'-3' nuclease activity but lacks a 3'-5' proofreading exonuclease activity. When fidelity is required, proofreading exonucleases such as Vent and Deep Vent (New England Biolabs) or Pfu (Stratagene) may be used (Cline *et al.* (1996), *Nucl Acids Res* 24:3456-3551). In another embodiment of the

invention, a single enzyme approach maybe used involving a DNA polymerase with intrinsic reverse transcriptase activity, such as *Thermus thermophilus* (Tth) polymerase (Bustin (2000), *J Mol Endo* 25:169-193). A skilled artisan may appreciate the variety of enzymes available for use in the present invention.

[0048] The methodologies and control gene primers of the present invention may be used, for example, in any kinetic RT-PCR methodology, including those that combine fluorescence techniques with instrumentation capable of combining amplification, detection and quantification (Orlando *et al.* (1998), *Clin Chem Lab Med* 36:255-269). The choice of instrumentation is particularly important in multiplex RT-PCR, wherein multiple primer sets are used to amplify multiple specific targets simultaneously. This requires simultaneous detection of multiple fluorescent dyes. Accurate quantitation while maintaining a broad dynamic range of sensitivity across mRNA levels is the focus of upcoming technologies, any of which are applicable for use in the present invention. Preferred instrumentation may be selected from the ABI Prism 7700 (Perkin-Elmer Applied Biosystems), the Lightcycler (Roche Molecular Biochemicals) and iCycler Thermal Cycler. Featured aspects of these products include high-throughput capacities or unique photodetection devices.

[0049] Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, practice the methods and use the control genes of the present invention. The following examples therefore, specifically point out the preferred embodiments of the present invention, and are not to be construed as limiting in any way the remainder of the disclosure.

EXAMPLES

Example 1: Selection of Control Genes

[0050] The control genes were selected by querying a Gene Logic rat tissue database to create expression profiles from a variety of rat cell and tissue samples.

[0051] This database was produced from data derived from screening various cell or tissue samples using an Affymetrix rat GeneChip® set. The rat cell and tissue samples that were analyzed include those that were not treated at all and that can be referred to as "normal," as they represent the laboratory rat population that has not been manipulated outside of normal daily activity within that setting. In general, tissue and cell samples were

processed following the Affymetrix GeneChip® Expression Analysis Manual. Frozen tissue or cells were ground to a powder using a Spex Certiprep 6800 Freezer Mill. Total RNA was extracted with Trizol (GibcoBRL), according to the manufacturer's protocol. The total RNA yield for each sample was 200-500 μg per 300 mg cells. mRNA was isolated using the Oligotex mRNA Midi kit (Qiagen) followed by ethanol precipitation. Double stranded cDNA was generated from mRNA using the SuperScript Choice system (GibcoBRL). First strand cDNA synthesis was primed with a T7-(dT24) oligonucleotide. The cDNA was phenol-chloroform extracted and ethanol precipitated to a final concentration of 1 μg/ml. From 2 μg of cDNA, cRNA was synthesized using Ambion's T7 MegaScript *in vitro* Transcription Kit.

[0052] To biotin label the cRNA, nucleotides Bio-11-CTP and Bio-16-UTP (Enzo Diagnostics) were added to the reaction. Following a 37°C incubation for six hours, impurities were removed from the labeled cRNA following the RNeasy Mini kit protocol (Qiagen). cRNA was fragmented (fragmentation buffer consisting of 200 mM Tris-acetate, pH 8.1, 500 mM KOAc, 150 mM MgOAc) for thirty-five minutes at 94°C. Following the Affymetrix protocol, 55 μg of fragmented cRNA was hybridized on an Affymetrix Rat Genome U34 array set for twenty-four hours at 60 rpm in a 45°C hybridization oven. The chips were washed and stained with Streptavidin Phycoerythrin (SAPE) (Molecular Probes) in Affymetrix fluidics stations. To amplify staining, SAPE solution was added twice with an anti-streptavidin biotinylated antibody (Vector Laboratories) staining step in between. Hybridization to the probe arrays was detected by fluorometric scanning (Hewlett Packard Gene Array Scanner). Following hybridization and scanning, the chips were analyzed for quality control, looking for major chip defects or abnormalities in hybridization signal. After the chips passed quality control, data were analyzed using Affymetrix GeneChip® version 3.0 and Expression Data Mining Tool (EDMT) software (version 1.0), S-Plus, and the GeneExpresss software system. Microarrays were scanned on a high photomultiplier tube (PMT) settings.

[0053] To prepare tissue samples from animals, e.g., rats, sterile instruments were used to sacrifice the animals, and fresh and sterile disposable instruments were used to collect tissues. Gloves were worn at all times when handling tissues or vials. All tissues were collected and frozen within approximately 5 minutes of the animal's death. The liver sections and kidneys were frozen within approximately 3-5 minutes of the animal's death. The time of

euthanasia, an interim time point at freezing of liver sections and kidneys, and time at completion of necropsy were recorded. Tissues were stored at approximately -80°C or preserved in 10% neutral buffered formalin.

[0054] Tissues were collected and processed as follows.

[0055] Liver

1. Right medial lobe - snap frozen in liquid nitrogen and stored at -80°C.

2. Left medial lobe - Preserved in 10% neutral-buffered formalin (NBF) and evaluated for gross and microscopic pathology.

3. Left lateral lobe - snap frozen in liquid nitrogen and stored at -80°C.

[0056] Heart- A sagittal cross-section containing portions of the two atria and of the two ventricles was preserved in 10% NBF. The remaining heart was frozen in liquid nitrogen and stored at - 80°C.

[0057] Kidneys (both)

1. Left - Hemi-dissected; half was preserved in 10% NBF and the remaining half was frozen in liquid nitrogen and stored at - 80°C.

2. Right - Hemi-dissected; half was preserved in 10% NBF and the remaining half was frozen in liquid nitrogen and stored at - 80°C.

[0058] Testes (both)- A sagittal cross-section of each testis was preserved in 10% NBF. The remaining testes were frozen together in liquid nitrogen and stored at - 80°C.

[0059] Brain (whole)- A cross-section of the cerebral hemispheres and of the diencephalon was preserved in 10% NBF, and the rest of the brain was frozen in liquid nitrogen and stored at - 80°C.

[0060] Gene expression data were then analyzed to identify those genes that were consistently expressed across a set of about 5,000 different tissue samples, *e.g.*, being called Present more than 95% of the time. For each of these samples, the mean average difference, standard deviation and CV were determined for each Affymetrix fragment on the rat U34 GeneChip®. The data were sorted by CV, and those gene fragments with values less than 40% were chosen for further analysis. Table 1 provides a list of approximately 858 genes with a coefficient of variation less than 0.44 and whose expression is considered not to vary across the normal and treated samples studied. For each gene listed, Table 1 also provides a GenBank Accession No., a Present frequency value, a mean expression level value and a coefficient of variation, expressed as CV. The GenBank Accession Nos. can be used to

locate the publicly available sequences, each of which is herein incorporated by reference in its entirety as of the priority date of this application (July 30, 2002).

Example 2: Quantitative PCR Analysis of Expression Levels using the Control Genes

[0061] The expression levels of one or more genes listed in Table 1 may be used to normalize gene expression data produced using quantitative PCR analysis. For example, the sequences may be used as Taqman probes, along with the forward and reverse primers for a gene in Table 1. Real time PCR detection may be accomplished by the use of the ABI PRISM 7700 Sequence Detection System. The 7700 measures the fluorescence intensity of the sample each cycle and is able to detect the presence of specific amplicons within the PCR reaction. The TaqMan® assay provided by Perkin Elmer may be used to assay quantities of RNA. The primers may be designed from each of the genes identified in Table 1 using Primer Express, a program developed by PE to efficiently find primers and probes for specific
sequences. These primers may be used in conjunction with SYBR green (Molecular Probes), a nonspecific double-stranded DNA dye, to measure the expression level of mRNA corresponding to the expression levels of each gene. This gene expression data may then be used to normalize gene expression data of other test genes.

[0062] Although the present invention has been described in detail with reference to examples above, it is understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims. All cited patents and publications referred to in this application are herein incorporated by reference in their entirety.

TABLE 1			
GenBank No.	Pres nt Frequency	Adjusted Mean	CV
NM_057141	0.9621	353.7302949	0.394573166
AA800364	0.9921	538.7477202	0.35144586
AA800501	0.9874	200.464431	0.271392863
AA801051	0.9946	594.9934288	0.327732429
AA801442	0.9937	472.4598982	0.314834757
AA848238	0.9516	341.660987	0.288832086
AA849262	0.9846	210.520306	0.37721446
AA944127	0.9522	178.105728	0.384011401
NM_031981	0.994	328.1689155	0.389666675
NM_031981	0.998	384.2180916	0.31917883
NM_019352	0.9964	577.4245502	0.233685612
AB008807	0.9906	375.3583161	0.389113636
NM_019213	0.9959	167.7105557	0.33117565
NM_031331	0.9928	299.5668687	0.388714827
NM_019191	0.9725	67.90177915	0.330844052
NM_053527	0.9608	151.0502046	0.316326745
AF003926	0.9947	331.151405	0.243708921
NM_031656	0.9624	70.8858116	0.349709856
NM_053553	0.9656	136.6785634	0.375165396
NM_053556	0.9816	193.2494812	0.365428046
NM_019201	0.9993	800.7665129	0.383779573
NM_022536	0.992	637.6998997	0.349019258
NM_031749	0.9833	260.969288	0.379520142
AF093139	0.9955	164.5811247	0.291070695
NM_053467	0.9967	645.032758	0.312688813
NM_019222	0.9959	241.1715455	0.306923163
NM_053707	0.9954	278.2968887	0.363812568
NM_057143	0.9969	491.5701144	0.377464693
NM_057141	0.99	334.7099186	0.372595428
NM_017284	0.9991	469.0115461	0.37489808
NM_053743	0.9986	599.9253754	0.318539549
NM_031603	0.9961	531.1713873	0.394886131
D37934	0.9794	212.2995269	0.267710366
NM_022598	0.9898	150.3146804	0.392695104
NM_022598	0.9978	410.4386698	0.355948912
NM_013076	0.9736	249.8757424	0.364628324
NM_019317	0.9734	83.51973777	0.373249306
NM_017236	0.9867	671.5891964	0.360897244
H32978	0.997	435.2398828	0.300833768
NM_031090	0.9619	70.31528575	0.399983405
NM_057209	0.9864	257.5044748	0.332226154
NM_012500	0.9895	150.6522809	0.302162035
K02816	0.9981	382.6421388	0.334260892
NM_022518	0.99	575.2287493	0.267122194
NM_031129	0.9986	672.686873	0.268407165
NM_012639	0.9592	157.9459425	0.307400434
NM_031974	0.9978	616.4278739	0.361748118
NM_013177	0.9988	787.1641147	0.381937146
NM_017101	0.9975	1067.896541	0.347227639
M57728	0.9729	108.3973358	0.368600327
AA684641	0.9692	132.9106769	0.312063584

TABLE 1			
G nBank No.	Present Frequency	Adjusted Mean	CV
AA799279	0.9991	839.3057142	0.325266583
AA799279	0.9947	568.1583462	0.347844769
AA799542	0.9924	273.5836187	0.370208871
AA799550	0.9973	470.9384047	0.370333288
AA799609	0.9912	134.1295318	0.334268614
AA799641	0.9966	276.4144125	0.307718893
AA799654	0.9981	296.1941725	0.351166278
AA799667	0.9908	248.9277967	0.291789627
AA799721	0.9629	114.4838534	0.373755794
AA799735	0.9644	126.8477716	0.292430382
AA799735	0.9813	137.5032487	0.318140687
AA799822	0.9906	162.7568631	0.360262563
NM_033096	0.9941	225.0546461	0.319505767
AA800015	0.9972	384.3536135	0.289608893
AA800039	0.9906	354.1901013	0.287620068
AA800053	0.9898	129.5213675	0.37530915
AA800170	0.9675	75.9629053	0.355273922
AA800198	0.9639	159.2105578	0.295644976
AA800210	0.9821	105.0330379	0.370992795
NM_013006	0.9898	237.3041636	0.391423327
AA800268	0.976	166.4768623	0.340868372
AA800651	0.9912	400.5374777	0.330167434
AA800669	0.9949	426.0164527	0.393889597
AA800787	0.9874	149.0104015	0.379600998
AA800814	0.9525	109.058537	0.389571008
AA801130	0.9957	263.9245532	0.38007823
AA801176	0.989	325.5564512	0.295890207
AA801230	0.9972	567.3071148	0.389999402
NM_032057	0.9955	179.3952918	0.296133732
AA817769	0.9941	185.0590031	0.323946319
AA817845	0.9951	380.1019029	0.242937824
NM_053682	0.9941	267.5028747	0.372604104
AA817892	0.9828	305.8321361	0.370745167
AA817907	0.9916	285.4664154	0.323851183
AA817945	0.9979	1077.847309	0.363411979
AA817967	0.9943	296.89826	0.323986488
AA818118	0.9951	324.9041145	0.349051053
AA818129	0.99	187.622771	0.301614267
NM_130405	0.9909	169.6695017	0.325635921
AA818203	0.9931	173.3889032	0.372430825
AA818246	0.9878	423.0700233	0.395146083
AA818534	0.9927	259.4059044	0.283803337
AA818568	0.9772	79.08353703	0.299639563
AA818669	0.9928	330.9312721	0.317467573
AA818697	0.9979	645.1875312	0.274054292
AA818778	0.9964	324.6127355	0.343055107
AA818788	0.988	128.4093671	0.374028119
NM_019907	0.993	178.4123338	0.361138088
AA819057	0.9974	559.6063582	0.246931544
AA819119	0.9621	91.25127707	0.380140187
AA819224	0.9812	135.2326929	0.383447682

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
NM_031745	0.9853	165.1417753	0.394406392
AA819318	0.9527	212.4831719	0.361885928
AA819362	0.9862	154.0922099	0.361381365
AA819364	0.9933	282.8186095	0.260938603
AA819367	0.991	135.7415399	0.34405043
AA819400	0.9886	135.3139207	0.345055159
AA819468	0.9986	320.3062492	0.334844865
AA819471	0.9678	102.5559907	0.342984629
AA819487	0.9736	138.4349905	0.345244474
AA819691	0.9941	431.8944648	0.382341107
AA819694	0.9714	103.4099525	0.372870205
AA819729	0.9931	289.9616028	0.32509288
AA819761	0.987	240.328863	0.353298909
AA819798	0.9961	412.8591164	0.38513809
AA819798	0.9977	543.3893727	0.333884343
AA848404	0.9812	470.973766	0.353247324
NM_133320	0.9901	374.0195578	0.370573918
AA848674	0.9709	198.5709918	0.268430905
AA848696	0.9584	113.9625472	0.393992922
AA848967	0.9551	371.1002902	0.378302744
AA849092	0.9941	265.2055344	0.361330207
AA849312	0.9896	261.6560729	0.302612314
AA849531	0.9977	452.3976272	0.258409286
AA849715	0.9955	427.4012934	0.324344933
AA849721	0.9759	429.6255416	0.3397144
AA849757	0.991	728.7550103	0.385766928
AA849766	0.9905	338.2837276	0.386484508
AA849767	0.9892	686.473722	0.370408924
AA849774	0.9543	199.0178366	0.290368511
AA849788	0.9899	277.0979159	0.302937581
AA849809	0.9847	213.0699948	0.395797677
AA849952	0.9836	249.6829894	0.317677787
AA849954	0.9523	86.97070267	0.337457286
AA849965	0.9908	209.8670776	0.346431239
AA850117	0.9611	228.5306101	0.38524339
AA850451	0.9922	427.5517004	0.379543087
AA850480	0.9932	408.6763631	0.386935606
AA850525	0.9702	335.2964034	0.260647679
AA850529	0.9771	383.9424729	0.372557101
AA850535	0.994	484.8463265	0.306355053
AA850550	0.9955	538.4247695	0.220155894
AA850569	0.9969	539.4661192	0.369801777
AA850624	0.966	114.0740815	0.388810835
AA850666	0.9933	372.2546801	0.294655078
AA850754	0.9731	100.307608	0.376838798
AA850894	0.9917	442.4665358	0.273072964
AA850907	0.9935	365.9563743	0.326461416
AA851161	0.9968	448.0118551	0.319626206
AA851202	0.9808	200.0256289	0.355469087
AA851214	0.9914	523.3549456	0.295434486
AA851251	0.996	296.8962387	0.303746

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AA851347	0.9982	438.4096815	0.299495219
AA851376	0.9982	499.552633	0.311653073
AA851397	0.9651	325.960527	0.365291903
AA851405	0.9773	114.6840333	0.327907652
AA851439	0.962	229.2705115	0.326726191
AA851464	0.9918	267.8901625	0.284156685
AA851641	0.979	179.8467866	0.376196046
NM_133324	0.9766	178.8689584	0.271247953
AA851686	0.9954	300.5452993	0.237099425
AA851701	0.9737	134.3344569	0.344149947
AA851728	0.9785	252.4667912	0.323521957
AA851739	0.9553	206.4328905	0.300390557
AA851765	0.97	828.6859018	0.305562042
AA851873	0.9721	329.8649568	0.359666828
AA851883	0.977	185.2628058	0.329071104
AA851909	0.9906	206.1724423	0.31357168
AA851920	0.9779	208.7672739	0.374968241
AA851938	0.9938	392.3178541	0.326446676
AA858457	0.9843	168.426778	0.309690174
NM_031153	0.9815	289.8076299	0.3760569
AA858551	0.9893	210.2360092	0.362121119
AA858660	0.9962	252.2957635	0.34707555
AA858718	0.9977	1028.984349	0.399132237
AA858833	0.9793	159.8825966	0.393184202
AA858867	0.9648	136.1726528	0.331092316
AA858990	0.9965	870.3627949	0.365647982
AA859100	0.9772	136.2773269	0.356567976
AA859201	0.9978	275.683128	0.306043339
AA859796	0.9534	71.01143176	0.389965008
AA859919	0.9881	142.7674515	0.333076793
AA859919	0.9988	667.1537331	0.292027993
AA866364	0.9931	138.1245174	0.38210287
AA866371	0.9633	142.6912204	0.320447204
NM_024394	0.9617	151.4019916	0.387696691
AA875431	0.9929	254.161646	0.360714718
AA875470	0.9605	291.3569793	0.273123402
AA875470	0.9683	127.7332542	0.328086924
AA875552	0.9933	196.7545027	0.275253333
AA875661	0.9853	74.47796632	0.335756096
NM_053739	0.9971	223.2643235	0.343236121
AA891546	0.972	65.30502026	0.399473194
AA891717	0.9819	135.7970398	0.291717192
AA891742	0.9844	120.1613409	0.361978153
AA891746	0.9946	310.8827948	0.393391849
AA891810	0.9808	328.0727833	0.30475391
AA891810	0.9858	152.5190759	0.339434397
AA891902	0.9702	51.59142223	0.349786011
AA891935	0.988	233.0024719	0.270449187
AA892120	0.9791	60.97800731	0.373214916
AA892313	0.9913	107.2159627	0.389027092
AA892394	0.9971	221.3345267	0.386151078

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AA892394	0.9952	129.1994687	0.397846098
AA892422	0.975	185.0194316	0.252043297
AA892505	0.9941	256.1674794	0.272080589
AA892550	0.9702	118.6361973	0.377480727
AA892789	0.986	236.3026727	0.297799259
AA892791	0.9855	175.6311243	0.325854158
AA892796	0.9973	621.5110927	0.313521756
AA892814	0.9959	421.7165288	0.32340475
AA893224	0.9918	129.2966135	0.319716751
AA893353	0.9939	289.2938276	0.343012426
AA893515	0.9965	268.1693134	0.293905313
AA893641	0.9655	124.631513	0.343387848
AA893683	0.983	87.44882196	0.347534606
AA893741	0.9921	192.2533724	0.269508199
AA893811	0.9736	84.60301216	0.398222929
AA894099	0.9813	285.142634	0.328324892
AA894101	0.9559	102.4471478	0.332265391
AA894101	0.9824	111.4560965	0.350773318
AA894131	0.9766	106.387669	0.383457001
AA894234	0.9841	236.1264994	0.284552291
AA894259	0.9966	449.0454763	0.312420033
AA899546	0.982	220.5971506	0.309976207
AA899672	0.9979	1656.075895	0.319960146
AA899691	0.9924	195.4072733	0.34078252
AA899743	0.9974	1071.973786	0.330354696
AA899911	0.9976	522.9454135	0.252669377
AA899959	0.9952	492.0446142	0.395266737
AA900078	0.9565	188.0606334	0.387115384
AA900156	0.9857	761.5282734	0.223294586
AA900187	0.998	471.5301948	0.257859064
AA900343	0.9773	209.1604636	0.322304252
AA900348	0.9502	212.7503091	0.340891055
AA900364	0.9652	162.5703442	0.284958346
AA900422	0.9604	404.2714995	0.356451649
AA900860	0.9869	167.3921533	0.383802956
AA900891	0.9524	149.0980669	0.315160786
AA900975	0.9978	868.1886108	0.37200699
AA901222	0.9956	480.1616103	0.309579403
AA901365	0.9995	890.2613004	0.32691327
AA923992	0.9576	114.2571622	0.396496378
AA923998	0.9659	225.9503235	0.280597467
AA924030	0.989	162.9127727	0.392729941
AA924079	0.9821	205.1877284	0.379629317
AA924092	0.9888	318.9907396	0.228510957
AA924169	0.9927	288.0750562	0.272550639
AA924317	0.9867	190.5012975	0.224003653
AA924339	0.999	1652.670033	0.282834255
AA924369	0.9936	365.6561614	0.261149514
AA924532	0.9632	344.493183	0.317913577
NM_031020	0.9604	62.91020969	0.397017329
AA924604	0.9938	318.2624476	0.354029504

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AA924609	0.9984	461.8011067	0.378571404
AA924654	0.9809	211.2829082	0.332834334
NM_053555	0.9507	311.2721659	0.318866213
AA924765	0.9972	324.8928559	0.224049302
AA924768	0.9896	599.7543673	0.309926225
AA924787	0.9896	481.0525157	0.333960186
AA924871	0.9647	148.8977646	0.393740956
AA925123	0.9984	850.7664442	0.264758138
AA925152	0.9946	699.0355291	0.239167455
AA925160	0.9823	190.1916631	0.373116063
AA925212	0.9973	611.8057286	0.294395182
AA925304	0.9902	302.2840247	0.272196777
AA925305	0.9837	411.7095514	0.285568444
AA925338	0.9803	298.8774464	0.290626881
AA925340	0.9991	521.8490052	0.283743847
AA925341	0.9669	241.6436392	0.360237623
AA925432	0.9735	225.7988151	0.350901777
AA925473	0.9959	622.4378044	0.399479916
AA925478	0.9819	341.7412759	0.372375386
AA925677	0.9608	201.0099572	0.335202855
AA925854	0.9842	201.3893423	0.31172844
AA925979	0.9878	458.165257	0.346902976
AA925983	0.999	599.5384168	0.391862852
AA926013	0.981	193.6558006	0.228102661
AA926098	0.9965	755.140021	0.247890637
AA926279	0.9703	258.1851509	0.330058893
AA926331	0.9658	142.9219708	0.389925982
AA933158	0.9771	163.6147359	0.267516634
AA942947	0.9693	175.6990963	0.375870721
AA943015	0.9726	170.0338035	0.360085756
AA943015	0.9858	388.2429204	0.341776907
AA943122	0.9762	419.8206702	0.320372184
AA943240	0.9636	203.3959179	0.34869085
AA943281	0.9691	305.0288964	0.378822839
NM_012913	0.9813	258.9236246	0.388849176
AA943421	0.9632	209.3195009	0.370589173
AA943500	0.9708	209.919368	0.281165988
AA943553	0.9872	519.9751425	0.382535341
AA943553	0.9966	665.5611215	0.379984839
AA943645	0.9933	581.0411338	0.381146596
AA943738	0.9859	137.0917646	0.271120535
AA943766	0.9957	400.0460991	0.370001453
NM_080909	0.9928	872.5315577	0.389923883
AA944203	0.9914	400.6244567	0.350853364
AA944335	0.9976	866.0289208	0.267386275
AA944347	0.9626	161.3067747	0.318180188
AA944445	0.9781	222.5978106	0.373397302
AA944451	0.9971	599.870325	0.323030108
AA944528	0.9954	425.8651494	0.22460926
AA944635	0.9801	322.8449619	0.312266733
AA944842	0.9861	299.6522749	0.237903089

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AA945089	0.9941	914.5995769	0.375028263
NM_133297	0.994	683.5483047	0.367321409
AA945740	0.9612	124.6495711	0.329660581
AA945746	0.9956	408.4484576	0.33612417
AA945746	0.9839	255.5524434	0.348157071
NM_131907	0.9962	370.0878264	0.339124904
AA945869	0.9792	179.2868431	0.327315212
AA946004	0.983	142.7065978	0.295987651
AA946018	0.9978	786.5592317	0.233292925
AA946038	0.9888	281.4038976	0.278216507
AA946205	0.9976	828.2592325	0.35788087
AA946432	0.9924	545.660241	0.259340094
AA946440	0.9982	570.5688594	0.245347075
AA955112	0.9962	291.1739164	0.212582952
AA955240	0.9973	647.7129713	0.274411246
NM_017359	0.996	231.4699675	0.329048264
AA955396	0.9879	243.1046885	0.255401546
AA955506	0.9877	242.1871379	0.383670644
AA955536	0.9823	187.9185985	0.269361535
AA956114	0.9955	143.3842105	0.398204182
AA956140	0.9938	776.6322184	0.395628638
AA956185	0.9823	250.4214737	0.341205084
AA956460	0.9955	399.3603811	0.336899504
AA956983	0.9853	295.1596861	0.34098573
AA956992	0.9928	417.4006281	0.269624667
AA956992	0.9976	491.7470975	0.243689773
AA957063	0.9941	391.7747852	0.319424296
AA957491	0.988	180.2576966	0.351832394
AA957649	0.9924	423.4676789	0.338824147
AA957676	0.9592	429.4318847	0.396293319
AA957777	0.9866	112.7434987	0.363054879
AA963072	0.9709	134.5270945	0.333884657
AA963094	0.9977	606.8333854	0.328724125
AA963170	0.987	118.5722127	0.275144443
AA963367	0.9977	817.7237717	0.369667036
AA963808	0.998	669.6077262	0.382254088
AA964054	0.9949	405.6577401	0.347892099
AA964064	0.9888	375.2686433	0.37526759
AA964082	0.9956	680.4697645	0.272604335
AA964114	0.9831	753.7315494	0.29001828
AA964362	0.9869	145.3535334	0.33101453
AA964366	0.9774	316.6869935	0.283982561
AA964607	0.9923	499.9287489	0.320740471
AA964624	0.9538	125.8056987	0.280619534
AA964630	0.993	389.0162941	0.283601586
AA964642	0.9755	372.5717109	0.358580387
AA965073	0.9802	626.0264683	0.341947106
AA996398	0.9501	142.9854577	0.356277705
AA996576	0.9856	411.2372292	0.313642878
AA996797	0.9889	366.872934	0.310438476
AA996939	0.994	369.2356692	0.339669675

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AA996974	0.9859	172.2669572	0.330914903
AA997052	0.9625	159.3897659	0.386386551
AA997184	0.9801	277.3537497	0.296759505
NM_053494	0.982	300.8210503	0.361139323
AA997929	0.9888	166.2636207	0.279152327
AA998158	0.9521	297.1376512	0.280727643
AA998435	0.9989	878.2922689	0.231761767
AA998523	0.9678	182.5520659	0.383305234
AA998556	0.9802	157.5036188	0.323268276
AA998843	0.9713	200.1415524	0.351313361
AA998893	0.9938	218.6451718	0.356607867
AI007743	0.9885	280.3187898	0.397692815
AI007750	0.9766	224.5812217	0.29763571
AI007920	0.9603	128.7892564	0.362125165
AI007987	0.9945	391.0597095	0.286018393
AI008372	0.996	586.5742499	0.29223874
AI008423	0.9933	197.7269351	0.251485242
AI008683	0.9969	590.7294525	0.250069145
AI008740	0.9606	182.2744427	0.352124819
AI008774	0.995	226.9116964	0.278990202
AI008784	0.9968	1245.190937	0.28202077
AI008931	0.9798	214.3402379	0.369067812
AI008958	0.9956	1352.486047	0.283940464
AI009079	0.9974	826.1352041	0.383871133
AI009157	0.9919	213.2822268	0.364389086
AI009200	0.9967	440.7137578	0.313553376
AI009350	0.9998	711.4169929	0.316543066
NM_053416	0.9988	2751.842839	0.354727984
AI009591	0.9601	115.0663701	0.297851506
AI009650	0.9861	290.3570616	0.313940175
AI009655	0.9884	408.8257028	0.386816933
AI009693	0.9945	704.5208126	0.38652766
AI009741	0.9983	502.0441942	0.347210566
AI009772	0.9982	661.4997157	0.357510104
AI009819	0.9833	385.974512	0.30650183
AI009936	0.9982	482.9852102	0.382801193
AI010034	0.9869	175.1325883	0.357119903
AI010342	0.9757	160.9745004	0.341899927
AI010362	0.9879	325.1416967	0.367946604
AI010422	0.9766	133.9096768	0.348370485
AI010452	0.9995	1364.399147	0.323158828
AI010518	0.9897	523.3835278	0.251025801
AI010758	0.9791	159.5785487	0.342965544
AI010944	0.9655	198.931889	0.348891534
AI011148	0.9888	509.7577822	0.297487627
AI011190	0.9848	244.2595909	0.263968956
AI011306	0.9735	186.9337855	0.299628693
AI011339	0.9878	339.3332871	0.295719531
AI011344	0.9985	808.9708186	0.242918026
AI011556	0.9933	307.1623657	0.361656015
AI011571	0.9913	231.1039866	0.349951837

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI011754	0.9601	357.2305561	0.259066046
AI011756	0.9845	505.6862363	0.335393314
AI012027	0.9905	498.9415909	0.330533351
AI012077	0.9685	183.5304607	0.28972612
AI012258	0.9627	232.9319168	0.397616077
AI012277	0.9815	224.0156956	0.315069042
AI012285	0.9969	455.2112947	0.251368311
AI012467	0.9755	247.9979552	0.393536674
AI012562	0.9819	247.0055648	0.399176309
AI012567	0.994	614.0183082	0.300419775
AI012636	0.9774	244.7954881	0.343753177
AI012641	0.9954	1107.58001	0.232410874
AI012937	0.9873	231.1042578	0.299231915
AI012947	0.9972	750.634375	0.34167501
AI012951	0.9969	589.2139983	0.348304745
AI013024	0.9941	972.6981993	0.377775478
AI013090	0.9848	294.8100632	0.307498244
AI013097	0.9949	470.59474	0.279198054
AI013204	0.9984	974.7028181	0.387451663
AI013350	0.9844	259.9148955	0.232030679
AI013363	0.9974	517.4391968	0.277623342
AI013555	0.9.	335.5652187	0.285524016
AI013564	0.96[illegible]	181.8970045	0.347588318
AI013697	0.9987	911.7514272	0.330567711
NM_031721	0.9825	322.6669671	0.384108704
AI013816	0.9654	635.7636797	0.31640618
AI013870	0.9903	267.8569007	0.324199211
AI013946	0.9985	532.8561833	0.216157143
AI014059	0.9806	391.8286644	0.306162721
AI028997	0.9851	246.3910601	0.293805901
AI029110	0.9941	292.5132162	0.268267155
AI029421	0.9787	293.3496299	0.204319564
AI029484	0.9977	367.1680693	0.227864083
AI029733	0.9987	1950.541312	0.296475897
AI029737	0.9916	520.438801	0.25582686
AI030147	0.9894	307.9855538	0.350690338
AI030192	0.9753	234.1326385	0.387161574
AI030248	0.9969	701.8459509	0.250558161
AI030430	0.9955	511.3544152	0.329914711
AI030751	0.9565	233.3235568	0.269114645
AI030799	0.9962	490.648901	0.386351868
AI030907	0.9939	322.801128	0.367771191
AI031035	0.9978	352.4611173	0.295253508
AI044112	0.9956	355.4596171	0.308803476
NM_053864	0.9816	329.9527469	0.325582005
AI044727	0.9941	399.1541549	0.24747723
NM_022595	0.9749	558.6476666	0.376778078
AI044863	0.9792	231.3400497	0.328631874
AI044872	0.9738	126.3552963	0.319073417
AI045003	0.9781	302.1675806	0.367003326
NM_053347	0.9792	425.6707106	0.262323344

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI045458	0.9987	834.2478872	0.281073528
AI045597	0.9795	236.7872339	0.362316181
AI045774	0.9984	597.5232402	0.361381584
AI045810	0.9758	119.8442776	0.37446961
AI058373	0.9714	166.5305651	0.310968224
AI058963	0.9795	226.9015804	0.318390827
AI058972	0.9639	163.4753435	0.314357814
AI059428	0.9631	337.5326409	0.343346037
AI059762	0.9628	101.8533271	0.350501189
AI060132	0.9923	869.1205218	0.296469343
AI060196	0.9987	652.5830242	0.247507825
AI060222	0.9954	274.2012257	0.299764113
AI070070	0.9762	237.8818411	0.32422811
AI070153	0.9524	241.7017398	0.274003857
AI070176	0.9983	508.310541	0.347627491
AI070399	0.9921	240.2738781	0.320172095
AI071243	0.9775	164.5641109	0.375047226
AI071773	0.9948	176.2687515	0.362438965
AI071946	0.9964	224.3185087	0.299033252
AI072081	0.9983	581.6999055	0.319003258
AI072121	0.9872	250.8923104	0.328781873
AI072555	0.9637	103.7773054	0.269538712
AI072666	0.9947	470.2436446	0.322438521
AI072675	0.9957	324.3957217	0.378467368
AI072885	0.9812	128.3180767	0.328993951
AI073030	0.997	529.5599398	0.226307567
AI073118	0.9816	132.8195789	0.330656304
AI073193	0.9983	387.0532434	0.314540172
AI073215	0.9879	422.9032397	0.327263785
AI073260	0.999	990.0658894	0.383685551
NM_053569	0.9896	203.3512567	0.226791767
AI101181	0.9741	97.17989413	0.28235925
AI101222	0.993	312.149685	0.31926142
AI101375	0.9933	415.4787877	0.391617213
AI101395	0.9725	160.7363746	0.388039479
AI101438	0.9709	64.45239629	0.38983282
AI101460	0.9974	399.5143745	0.355650986
AI101659	0.9988	627.0523045	0.329943
AI101864	0.9983	1112.010461	0.276584797
AI101934	0.973	160.6040766	0.331698393
AI102046	0.9785	187.0808649	0.28794593
AI102080	0.9882	291.239455	0.268077727
AI102191	0.9768	160.2239891	0.276862462
AI102252	0.9936	186.9971017	0.289805898
NM_053436	0.973	227.2297147	0.385866717
AI102438	0.9956	321.3897237	0.394869448
AI102612	0.9975	571.3800141	0.289213412
AI102734	0.9938	530.6640201	0.385875553
AI102935	0.9884	324.8164874	0.3850101
AI102978	0.9903	147.7101205	0.381554081
AI102991	0.998	389.6494934	0.211087424

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI103094	0.9979	873.9486249	0.282128391
AI103129	0.9774	740.612351	0.322658411
NM_031146	0.9852	438.0155262	0.279684593
AI103377	0.9844	226.9637436	0.305656574
AI103379	0.9981	528.362849	0.191923868
AI103428	0.9832	141.4527273	0.375933874
AI103521	0.9907	384.7064269	0.310270528
AI103717	0.95	277.674925	0.301083424
AI103718	0.998	991.3056425	0.220129724
AI103848	0.9836	146.2910591	0.304856826
AI103950	0.9684	85.7813618	0.38997044
AI103954	0.9965	345.2301178	0.370824066
AI104231	0.9752	132.4494652	0.325845881
AI104234	0.9786	374.7482362	0.356449801
AI104239	0.979	125.3204702	0.332461387
AI104247	0.9691	212.5064712	0.340422775
AI104250	0.9641	219.0354755	0.325898084
AI104283	0.9922	289.7160716	0.266096751
AI104320	0.9906	303.2373949	0.248088041
AI104388	0.9505	156.6673508	0.329438846
AI104488	0.9672	117.8088465	0.229489312
AI104536	0.9986	1025.536272	0.284465579
NM_022518	0.9936	711.8999891	0.304763664
AI104600	0.9521	122.2863459	0.349242431
AI104753	0.9524	326.7032346	0.399036819
AI104864	0.9868	376.8258789	0.356488841
AI104878	0.9972	438.505944	0.358693351
AI104914	0.9956	199.5044251	0.268591505
NM_080781	0.9663	107.7513281	0.335662126
AI105072	0.9972	395.9783073	0.387243663
NM_057205	0.9978	283.3131956	0.293441255
AI105087	0.9933	515.1104067	0.352504039
AI105149	0.9983	911.5392665	0.263156658
AI105265	0.9538	217.2837741	0.341799777
AI105345	0.9861	155.8460745	0.364213518
AI105352	0.9938	141.9075167	0.361489718
AI105431	0.998	356.6841375	0.378020827
AI111683	0.9915	184.2053628	0.241117848
AI111975	0.999	192.3560148	0.3351647
AI112092	0.9954	269.3073934	0.349380313
AI112250	0.9968	653.2935303	0.378593708
AI112512	0.9598	75.40270266	0.386823108
AI113020	0.9844	232.0838805	0.311528728
AI136231	0.9537	132.4605103	0.376730451
AI136564	0.9761	278.6318378	0.35005281
AI136669	0.9958	518.5194087	0.351974463
AI137232	0.9988	469.6158446	0.366380187
AI137298	0.9923	230.6509496	0.270580577
AI137582	0.9799	135.3498294	0.397497765
AI138002	0.9942	219.6829104	0.24859447
AI144657	0.9923	129.4479951	0.312677319

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI144668	0.9909	297.1273683	0.314893774
AI144956	0.9904	207.7546316	0.264521312
AI145332	0.9538	129.0242513	0.37788731
AI145362	0.9719	120.3802137	0.339227369
AI145368	0.9917	254.051109	0.364636436
AI145614	0.9969	441.5437287	0.356577697
AI145627	0.9917	327.5608757	0.302478299
AI145853	0.9823	133.5529582	0.319436187
AI146034	0.9925	154.6275229	0.324965874
AI146037	0.9941	168.134647	0.275451237
AI146090	0.9967	305.2987201	0.284411247
AI146170	0.9944	210.5029925	0.228534495
AI168933	0.9927	219.3020558	0.278034578
AI168950	0.99	318.010511	0.362293659
AI168974	0.9754	214.4003991	0.377276222
AI168979	0.9801	212.2729809	0.331980427
AI168986	0.9746	156.7979716	0.360198438
AI169063	0.9964	290.6780252	0.385981295
AI169154	0.9968	336.5242284	0.307023873
AI169170	0.9979	769.7878541	0.398738752
AI169269	0.977	137.357114	0.35503002
AI169272	0.9696	80.83140252	0.339825829
AI169343	0.9727	166.989764	0.268576719
AI169377	0.9889	180.0298019	0.390844085
AI169461	0.9973	866.2081039	0.336846289
AI169611	0.9986	503.4638109	0.36365031
AI169615	0.9985	663.3604215	0.326765274
AI169641	0.9962	363.828376	0.277445228
AI169642	0.9856	143.4258646	0.275093398
AI170247	0.9568	102.1625518	0.31850578
AI170265	0.9961	361.1879451	0.39819102
AI170357	0.9719	133.7080641	0.281598384
AI170388	0.9935	162.5694081	0.354744306
AI170400	0.9508	75.04534008	0.377930524
AI170414	0.9824	281.1240432	0.292176861
AI170532	0.9979	325.7623378	0.256357803
AI170663	0.9919	340.0625768	0.39841173
NM_032079	0.9912	212.7965304	0.383477936
AI170780	0.9978	403.7354889	0.31491612
AI170797	0.9898	362.0104956	0.367765173
AI170807	0.9943	244.3697528	0.250841845
AI170821	0.9835	115.6515135	0.35887866
AI171212	0.9978	775.9022683	0.275974842
AI171230	0.9719	69.49621762	0.348752498
AI171232	0.9996	746.0904006	0.390223181
AI171272	0.9961	584.7944874	0.273294529
AI171273	0.9838	409.6569296	0.341742937
AI171314	0.9894	690.4176735	0.399646269
AI171345	0.9857	121.9008899	0.300431813
NM_030836	0.9942	222.0517603	0.339780512
AI171561	0.9974	913.0878863	0.23190465

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
NM_019208	0.9812	125.3500482	0.338940828
AI171661	0.9675	108.2601624	0.280616401
AI171737	0.9904	251.8042864	0.37508985
AI171764	0.9973	487.4162473	0.277969318
AI171768	0.9941	333.6968205	0.399552284
AI171781	0.9916	195.2329285	0.325496907
AI171783	0.9935	277.3722053	0.390225646
AI171798	0.9511	96.82212997	0.357869848
AI171809	0.9786	121.0932339	0.375796802
AI171870	0.9849	205.0661444	0.333133061
AI171882	0.9965	253.5176312	0.301825594
AI171951	0.991	200.0156482	0.247113526
AI171952	0.9979	575.4556191	0.295443738
AI171953	0.9927	553.6106997	0.357140612
AI172001	0.982	118.9182618	0.358781789
AI172069	0.9579	55.27189598	0.301195
AI172074	0.9837	135.6336179	0.35943329
AI172092	0.9622	108.3322689	0.317645185
AI172105	0.9964	431.8804655	0.3638466
AI172106	0.9559	84.1857301	0.340208075
AI172196	0.9848	219.3575094	0.331715935
AI172214	0.9946	416.072214	0.309679658
AI172218	0.9678	136.6434257	0.298583323
AI172301	0.9895	280.4677498	0.327001975
AI172358	0.9609	229.83719	0.287010264
AI172472	0.9882	178.8898637	0.356223766
AI172537	0.9762	126.3743411	0.365833038
AI175001	0.9659	61.52159827	0.398114591
AI175008	0.9927	259.7040826	0.362558835
AI175044	0.9575	219.3801203	0.389920735
AI175266	0.9973	335.3095311	0.26393186
AI175366	0.9878	219.4067753	0.316098431
AI175467	0.9974	1050.953111	0.364080843
AI175477	0.9975	658.9995781	0.339519262
AI175512	0.999	1013.050673	0.248961012
AI175547	0.9599	86.61951632	0.316920617
NM_053969	0.9975	342.207506	0.23220273
AI175991	0.9735	93.66991174	0.324717316
AI176016	0.9896	118.3407824	0.35524637
AI176121	0.9984	1070.60159	0.328798698
AI176140	0.9985	1167.568018	0.301694674
AI176304	0.9927	123.8167239	0.335661085
AI176308	0.9965	366.1948025	0.338165832
AI176309	0.9542	86.00737984	0.344481111
AI176356	0.9946	109.3821659	0.389383607
AI176401	0.9844	124.7746569	0.350696016
AI176420	0.9925	201.397161	0.350564698
AI176491	0.9919	403.6217364	0.372574341
AI176511	0.9689	113.8307692	0.39779294
AI176581	0.9974	319.3364659	0.297959615
NM_031603	0.9949	216.3561619	0.362669512

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI176680	0.9875	447.7928097	0.319707122
AI176700	0.9947	219.7853067	0.396439967
AI176724	0.9903	209.9725455	0.302455154
AI177025	0.998	610.2210784	0.281843657
AI177104	0.9826	112.3718013	0.36436637
NM_130823	0.9867	1029.21364	0.382011417
AI177275	0.9552	151.4979672	0.387740506
AI177285	0.992	464.7912768	0.382919686
NM_053323	0.9988	1040.320182	0.36489234
AI177491	0.9963	259.4459705	0.30966825
AI177513	0.9925	309.6226866	0.340314119
AI177590	0.9662	136.1993835	0.305817502
AI177593	0.9972	754.5478523	0.336073841
NM_053798	0.9932	228.7776568	0.379213177
NM_022593	0.9964	231.7472596	0.352931313
AI177765	0.9947	242.1444179	0.376897727
AI177866	0.9944	235.3041612	0.359760014
AI177871	0.9978	493.0204781	0.36528958
AI177873	0.9732	144.5782393	0.323229199
AI177875	0.9749	169.1441977	0.327244948
AI177894	0.9978	381.7493132	0.277235768
AI177902	0.978	266.8474397	0.364559195
AI177919	0.9746	156.4655233	0.307400879
AI177921	0.9989	357.8900752	0.231216519
AI178052	0.9942	210.9919805	0.3269212
AI178239	0.9946	593.0948035	0.316485994
AI178378	0.974	113.3597499	0.35910838
AI178441	0.9693	123.977961	0.3713985
AI178503	0.9805	161.8575386	0.330391311
AI178526	0.9886	237.4170053	0.351527825
AI178644	0.9698	137.6729967	0.319376532
AI178763	0.9953	470.4968798	0.293156364
AI178830	0.9803	224.383254	0.374635776
AI178955	0.9978	647.2812159	0.338554719
AI179239	0.992	158.9663152	0.393554903
AI179243	0.9584	88.44774693	0.35176644
AI179327	0.9979	769.0504848	0.342140031
AI179329	0.9616	154.42071	0.265469568
AI179335	0.999	516.3069202	0.397506405
AI179355	0.9974	440.0164012	0.302809917
AI179356	0.999	561.1786991	0.297285533
AI179380	0.9927	471.0344443	0.399527454
AI179478	0.9899	388.0292776	0.311100554
AI179587	0.9609	181.1107877	0.27873237
AI179620	0.961	115.4729915	0.386630951
AI179636	0.9952	340.9861432	0.253360334
AI179640	0.9733	101.3470166	0.317864614
AI179711	0.9917	161.2168747	0.308572322
AI179833	0.9978	601.0236764	0.205199054
AI179840	0.972	274.1603007	0.324552252
AI179865	0.9841	437.9356753	0.284891811

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI179901	0.9957	309.429126	0.297663083
AI180015	0.9994	614.8581658	0.333863576
AI180081	0.9738	389.8384712	0.311918656
AI180108	*0.9864*	*284.6340916*	*0.330955649*
AI180224	0.9959	277.0615562	0.26693795
AI180259	0.9973	740.007384	0.269317518
AI180283	0.9766	336.6624044	0.383017026
AI180400	0.9968	483.0335627	0.38594082
NM_133324	0.964	108.365682	0.321158744
AI180426	0.9917	193.4211032	0.399608831
AI180441	0.9793	177.7127788	0.221639307
AI227612	0.9872	131.411593	0.379405748
AI227705	0.9973	373.9045536	0.308933462
AI227743	0.99	197.5316719	0.39445111
AI227884	*0.9981*	*1267.180243*	*0.223517344*
AI227887	0.9987	690.3196835	0.344995772
AI227894	0.9914	150.3145056	0.266321451
AI227962	0.9693	138.7234968	0.332020291
AI228112	0.9991	577.493851	0.329771829
AI228165	0.981	245.9051905	0.288970498
AI228249	0.9917	429.499532	0.295305029
AI228383	0.9684	118.3906252	0.308243873
AI228455	0.9592	252.5491309	0.259258531
AI228582	0.9931	244.0781278	0.299501991
AI229104	0.9973	418.4519495	0.234220274
AI229251	*0.9981*	*1138.337459*	*0.262304468*
AI229441	0.9967	720.1847476	0.286485755
AI229487	0.9972	326.6584951	0.278494869
AI229595	0.9949	334.9399022	0.360754811
AI229702	0.9886	220.6531992	0.314985308
NM_031342	0.9864	412.8077837	0.286440425
AI230069	0.9884	252.0236987	0.315987791
AI230073	0.9973	396.2082614	0.258575264
AI230192	0.9968	592.2203167	0.261516543
AI230248	0.9949	420.0225797	0.337867921
AI230278	0.9967	324.0160367	0.337131197
AI230308	*0.9803*	*180.5401476*	*0.350570361*
AI230503	0.9844	135.0925865	0.332488962
AI230635	0.9949	280.9665814	0.247155413
AI230778	0.9804	107.5929071	0.343369569
AI230912	0.9954	200.5872543	0.353036114
AI231017	0.9914	198.300742	0.381906854
AI231038	0.9956	250.0682523	0.277155064
AI231050	0.9943	410.8050546	0.253822599
AI231071	0.997	393.0335939	0.198604907
AI231201	0.9983	408.0126423	0.261904403
AI231471	0.9956	346.9078164	0.371698402
AI231491	0.9912	191.2310661	0.37822703
AI231773	0.9964	604.7876854	0.27563454
AI231785	0.9978	823.1725047	0.304029365
AI231812	0.982	210.7545364	0.282818931

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI231886	0.9978	443.3205068	0.368330282
AI232030	0.9805	402.2272212	0.353433208
AI232033	0.9926	258.5749225	0.304872148
AI232060	0.9826	129.6564971	0.320998742
AI232101	0.9941	610.1122963	0.275151496
AI232112	0.973	208.4695725	0.342289609
AI232129	0.9874	161.8677131	0.257738135
AI232159	0.9844	248.331737	0.349563487
AI232163	0.9791	499.4724254	0.367022262
NM_030586	0.9845	162.7074577	0.354117606
AI232274	0.9944	229.1965947	0.309473807
AI232296	0.9575	375.5156737	0.332664579
AI232321	0.9765	95.62859761	0.357723537
AI232354	0.9963	322.4466506	0.305484765
AI232510	0.9636	204.8272079	0.384522622
AI232639	0.953	114.8533235	0.370466228
AI232731	0.9661	207.339048	0.371840978
AI232734	0.9981	379.6275284	0.307581158
AI232800	0.9504	193.9482279	0.347453565
AI232807	0.983	197.1120336	0.309983248
AI232841	0.9903	307.2566121	0.312565038
AI232887	0.9942	204.7572949	0.363005514
AI232974	0.9922	259.9191333	0.365849096
AI232979	0.9901	232.304106	0.320933431
AI233096	0.9573	188.669731	0.368765567
AI233204	0.9956	1010.090755	0.339942787
AI233222	0.9993	768.2022698	0.307770797
AI233267	0.9768	115.1184504	0.370933612
AI233308	0.9935	151.0816592	0.349681314
AI233316	0.9941	301.9356829	0.36245711
AI233350	0.9919	228.184242	0.363111901
AI233370	0.9859	189.3310194	0.376119729
AI233698	0.9969	198.6385487	0.322074038
AI233728	0.9612	143.8504827	0.392284441
AI233915	0.9968	440.8259317	0.348972124
AI234008	0.9763	144.3922001	0.323216362
AI234040	0.9959	214.889063	0.282330926
AI234149	0.9894	147.2986378	0.346303317
AI234223	0.9943	155.5855792	0.295689739
AI234237	0.9898	128.0604113	0.365310901
AI234292	0.9666	132.0303364	0.371450337
AI234336	0.9606	108.0872625	0.348978236
AI234872	0.9933	342.8342984	0.348857143
AI234933	0.9735	437.5597637	0.362073729
AI235054	0.9805	158.1214142	0.341417567
AI235219	0.9903	372.8995033	0.398319082
AI235238	0.9975	828.1063382	0.269155653
AI235271	0.9859	210.1674784	0.269465284
AI235397	0.9937	263.135593	0.33163326
AI235403	0.9927	295.6660806	0.249674383
AI235502	0.9674	227.3319345	0.366379508

TABLE 1			
GenBank No.	Present Frequency	Adjusted Mean	CV
AI235508	0.9741	166.283959	0.274104484
AI235510	0.9981	1041.871028	0.289468985
NM_022518	0.9893	485.7282713	0.364274933
AI235885	0.9861	143.7381502	0.330077747
AI235901	0.9788	116.0377302	0.33953459
AI235962	0.9923	170.256446	0.228940909
AI236003	0.9911	116.7852026	0.368833607
AI236307	0.9931	647.0745083	0.3535376
AI236318	0.9905	145.0224218	0.368534279
AI236520	0.9859	230.0384121	0.310186585
AI236523	0.9693	79.93762767	0.334615459
AI236529	0.9893	267.9688613	0.215091202
AI236570	0.9972	1503.592959	0.295288772
AI236681	0.9979	434.0489709	0.388139121
AI236691	0.9938	329.1311041	0.374653742
AI236704	0.9847	87.82502754	0.361266423
AI236745	0.9936	232.0804362	0.235151157
AI236763	0.9736	114.0971841	0.323485716
AI236783	0.9988	405.5882713	0.270457401
AI236800	0.9588	130.7285204	0.364611624
AI236928	0.9889	249.3895955	0.311710968
AI237199	0.9505	96.75330112	0.385809363
AI237311	0.9975	994.9034091	0.307945865
NM_053989	0.9855	152.3314711	0.348698125
AI237700	0.9899	259.4171499	0.318929992
NM_031326	0.994	181.3514518	0.358704788
AI237861	0.9915	252.7434616	0.257047104
AI237872	0.9856	177.6381287	0.287293468
AI639425	0.9834	69.09078765	0.309131529
NM_057097	0.9897	196.9213407	0.392392697
S70803	0.9906	176.564558	0.296587753
NM_022588	0.9586	73.308782	0.367974984
NM_013221	0.9839	101.6691063	0.364721399
NM_022595	0.9955	303.915792	0.34309984
NM_053799	0.9948	362.974216	0.304148568
U53859	0.9911	598.5976337	0.357330309
NM_013050	0.9878	220.0914437	0.370223521
NM_053331	0.9996	556.6565158	0.26197082
U75392	0.9967	514.1769739	0.263076873
NM_021765	0.975	119.8855262	0.279960896
NM_017276	0.9834	371.4916806	0.349680389
Y13336	0.9959	552.6661681	0.270873633